WWW用インテリジェントプロトコルアナライザ

3F－2

大岸　智彦　井戸上　彰　加藤　聰彦　鈴木　健二
国際電信電話株式会社　研究所

1．はじめに

インターネットの普及により TCP/IP に従う通信が広く行われている。それに伴い、スループット低下時などに TCP/IP 通信を詳細に解析することが必要となっている。筆者らはこれまでに、回線上を流れる情報をモニタし、IP/TCP に従ったフォーマット解析を行うとともに、通信システムの TCP の振舞いを推定する TCP 用インテリジェントプロトコルアナライザ[1]を開発した。しかしながら、通信システム全体の動作をより明確化するためには、TCP だけでなく、アプリケーションレベルのプロトコルまでを含めた解析が必要である。そこで筆者らは、WWW を対象として、HTTP (Hypertext Transfer Protocol)[2]や HTML (Hypertext Markup Language)[3]までを含めて通信システムの動作を詳細に解析する WWW 用インテリジェントプロトコルアナライザを開発している。本稿ではその概要について述べる。

2．HTTP/HTML の概要

HTTP は、WWW で使用されるプロトコルであり、クライアント側は、リクエスト（REQ）により、テキストや画像などの情報を要求し、サーバ側はレスポンス（RSP）により要求された情報を転送する。REQ と RSP のフォーマットは、パラメータを格納するヘッダ部とテキスト／画像等を格納するコンテント部に分かれる。REQ には、要求する情報の所在を示す URL (Uniform Resource Locator) やブラウザが表示可能なファイル形式などのパラメータが含まれる。RSP には、コンテントのトータルバイト数、コンテントタイプなどのパラメータとともに、コンテントが含まれる。コンテントサイズが大きい場合は、複数の TCP セグメントに分けて転送される。

HTML は、WWW で使用されるテキスト形式のコンテントを規定するための言語であり、参照するコンテントの URL を指定するリンク情報の記述を可能とする。リンク情報は、ドキュメント内の画像情報などのように、ブラウザが自動的にアクセスする能動的リンクと、ハイパーリンクのように、ユーザが操作した時点で参照されたコンテントへのアクセスが行われる受動的リンクとに分かれる。

Intelligent Protocol Analyzer for WWW
Tomohiko Ogishi, Akira Idoue, Toshihiko Kato and Kenji Suzuki
KDD R&D Laboratories

3．機能

本アナライザは以下の機能を実現する。

(1) TCP の解析機能に加え、以下のような HTTP の解析機能を有する。

・HTTP に従ったフォーマット解析を行う。

・TCP-HTTP 間のインタフェースを明確化し、HTTP 用の状態／内部変数を導入することにより、REQ/RSP の対応づけや、リクエスト中止の検出など、HTTP の振舞いを推定する。

(2) HTTP により転送されたデータが HTML テキストの場合は、リンク情報を抽出し、後に発行されたリクエストとの対応づけを行う。この際、能動的リンクについては一連の情報要求として扱い、受動的リンクについてはユーザのアクセス時に別の情報要求が行われたとみなす。

4．設計

4.1．概要

図1に示すように、本アナライザは、IP、TCP、HTTP のプロトコルの解析を行う。TCP では、対象とする通信システムを指定し、そのシステム及びそのシステムと通信する全ての通信システムに関し、コネクション毎に状態／内部変数の推定を行う。TCP のユーザデータは、TCP の状態遷移の中で、プリミティブとして HTTP エミュレーション部に通知される。HTTP エミュレーション部では、TCP コネクション毎に、HTTP のコネクションを管理し状態／内部変数の推定を行う。

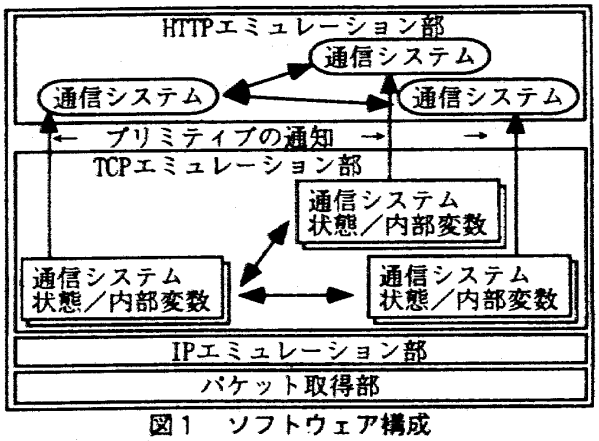


図1　ソフトウェア構成

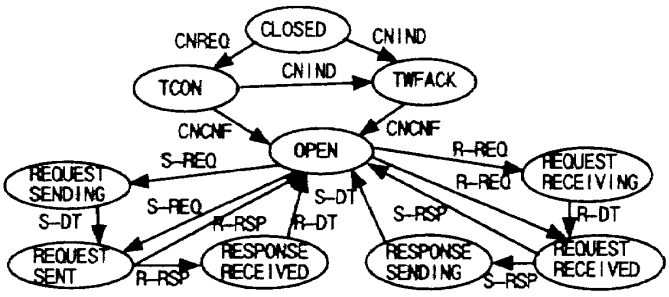
4.2．TCP プリミティブの通知

TCP エミュレーション部は、TCP プリミティブと呼ぶパラメータを用いて HTTP エミュレーション部でのイベントを通知する。TCP プリミティブとして

は、コネクション確立／解放などのTCPでの状態遷移を通知するものと、HTTPでのデータ転送を通知するものとを規定した。前者としては、コネクション確立要求（CNREQ）、コネクション確立指示（CNIND）、コネクション確立完了（CNCNF）、コネクション解放（CLOSE）を、後者としては、データ送信（DTSND）、データ受信（DTRCV）をそれぞれ定義した。なお、これらのプリミティブは、TCPの振舞いの推定機能により、HTTPにおいて処理されるタイミングで正しくHTTPエミュレーション部に通知される。

4.3. HTTPの振舞いの推定

HTTPエミュレーション部では、TCPコネクション毎に、本アナライザが独自に定義した状態／内部変数を推定する。図2に、正常シーケンスのみを抜粋したHTTPの状態遷移図を示す。HTTPの状態は、TCPコネクションの状態、REQ/RSPの送受信、コンテントの転送の途中／完了に分けて識別する。内部変数には、コンテント転送バイト数、コンテントタイプなどがあり、これらはTCPプリミティブやHTTPのヘッダ部の情報をもとに更新する。



S-REQ: REQを含むDTSNDプリミティブ、S-RSP: RSPを含むDTSNDプリミティブ、
S-DT: コンテントを含むDTSNDプリミティブ、R-REQ: REQを含むDTRCVプリミティブ、
R-RSP: RSPを含むDTRCVプリミティブ、R-DT: コンテントを含むDTRCVプリミティブ
(注)全ての状態において、CLOSEプリミティブにより、CLOSEDに遷移する。

図2　HTTPの状態遷移図（正常シーケンスのみ抜粋）

4.4. リンク情報の解析

HTTPエミュレーション部では、コンテントがHTMLテキストの場合は、リンク情報を抽出する。リンク情報の抽出及び能動／受動の識別は、<IMG SRC="URL"> 等のキーワードにより判断する。また、リンク情報の対応づけを行うため、各REQに対し、個別のセッションID（SID）を割り当てる。図3に示すように、HTTPエミュレーション部は、リンク情報をリストとして保持する。リンクリストは、URL、SID、親リンクを示すRSID（Root SID）から構成される。能動的リンクは、現在のセッションの子供のセッションと考え、子SIDを付加する。REQを解析する度、リンク情報がリストに存在するか否かを検査し、受動的リンクの場合（SIDが未定の場合）やリンクリストに存在しないURLの場合は、新しいSIDを割り当てる。

以上のように、各REQに対し、SID、RSIDなどの番号を付加することにより、要求されたテキストや画像の対応づけを解析することができる。

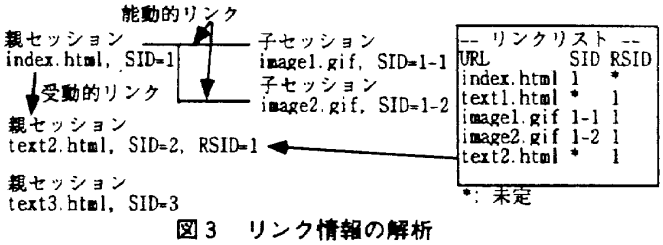


図3　リンク情報の解析

4.5. 解析例

図4にHTTPのエミュレーションの解析例を示す。これは、クライアント側のシステムのみに着目して、TCP及びHTTPの振舞いの推定を行った場合の解析例である。TCPコネクション確立／解放時には、HTTPは、状態遷移のみを表示する。REQ/RSPに関しては、そのパラメータを表示するとともに、SID等の内部変数を表示する。コンテントに関しては、そのTCPセグメントに含まれるオフセット範囲とコンテントタイプを表示する。コンテントがHTMLテキストの場合は、リンク情報を表示する。

```
21:46:15.687002 SYN->    SrcIP=*.*.*.* SrcP=1453 DstIP=paul DstP=80 CTL=S ...
                         [STATE=SYN-SENT ....]
                         *** Starting slow start... ***
                         [HTTP: STATE=TCON CID=1]
21:46:15.687324 SYN,ACK<- SrcIP=paul SrcP=80 DstIP=*.*.*.* DstP=1453 CTL=SA....
21:46:15.688610 ACK->    SrcIP=*.*.*.* SrcP=1453 DstIP=paul DstP=80 CTL=A ...
                         [STATE=ESTABLISHED ....]
                         [HTTP: STATE=OPEN CID=1]
21:46:15.695463 DT->     SrcIP=*.*.*.* SrcP=1453 DstIP=paul DstP=80 CTL=PA ...
                         [STATE=ESTABLISHED ....]
                HTTP->   GET / HTTP/1.0
                         Host: www.kdd.co.jp Accept: image/gif, image/jpeg
                         [HTTP: STATE=REQUEST_SENT CID=1 SID=1]
                         ### HTTP-Request completely sent ###
21:46:15.699491 DT<-     SrcIP=paul SrcP=80 DstIP=*.*.*.* DstP=1453 CTL=A....
                         [STATE=ESTABLISHED ....]
                HTTP<-   HTTP/1.1 200 OK
                         Content-type: text/html
                         [HTTP: STATE=REQUEST_SENDING CID=1 TYPE=text/html
                         BYTES=0]
21:46:15.700136 DT<-     SrcIP=paul SrcP=80 DstIP=*.*.*.* DstP=1453 CTL=A....
                         [STATE=ESTABLISHED ....]
                HTTP<-   ---- text/html (1-1460) ----
                         [HTTP: STATE=REQUEST_RECEIVING CID=1 TYPE=text/html
                         BYTES=1460]
21:46:15.700301 DT<-     SrcIP=paul SrcP=80 DstIP=*.*.*.* DstP=1453 CTL=PA....
                         [STATE=ESTABLISHED ....]
                HTTP<-   ---- text/html (1460-2256) ----
                         [HTTP: STATE=OPEN CID=1 TYPE=text/html BYTES=2256]
                HTML<-   [LINK_ID=1
                         ACTIVE_LINK=/graphics/sample.gif (1-1), ...(1-2)
                         PASSIVE_LINK=
                         http://www.lab.kdd.co.jp/, ...]
                         ### HTTP-Request completely received ###
21:46:15.700361 FIN<-    SrcIP=paul SrcP=80 DstIP=*.*.*.* DstP=1453 CTL=FA....
                         [STATE=ESTABLISHED ....]
                         [HTTP: STATE=CLOSED CID=1]
                         ### HTTP connection normally closed ###
```

(注) CID: HTTPのコネクションID

図4　解析例（クライアント側のみ）

5．おわりに

本稿では、HTTPやHTMLを含めて通信システムの動作を解析するWWW用インテリジェントプロトコルアナライザについて述べた。最後に日頃ご指導頂くKDD研究所村上取締役所長に感謝する。

参考文献

[1] T.Kato et.al., "Intelligent Protocol Analyzer with TCP Behavior Emulation for Interoperability Testing of TCP/IP Protocols," Proc. of FORTE/PSTV'97, pp.449-464, Nov.1997
[2] T.Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.0.," RFC1945, May 1996
[3] T.Berners-Lee et.al., "Hypertext Markup Language - 2.0.," RFC1866, Nov.1995